NEUROMANAGEMENT

People and Organizations

Edited by Michela Balconi

LED Edizioni Universitarie di Lettere Economia Diritto

IrcCAN
International Research Center for Cognitive Applied Neuroscience

ISSN 2724-3540
ISBN 978-88-7916-952-3

LED Edizioni Universitarie di Lettere Economia Diritto
Via Cervignano 4 - 20137 Milano
https://www.lededizioni.com
https://www.ledonline.it/neuropsychologicaltrends/

Printed by Logo

Index

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/

Big Data: neuroscience at the service of management

Federico Cassioli [1,2] **- Michela Balconi** [1,2]

[1] *International Research Center for Cognitive Applied Neuroscience (IrcCAN), Catholic University of the Sacred Heart, Milan, Italy*

[2] *Research Unit in Affective and Social Neuroscience, Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy*

DOI: https://dx.doi.org/10.7359/952-2020-cass federico.cassioli@unicatt.it

1. Introduction

William Edwards Deming, an American engineer and pioneer of quality management, is credited with the iconic phrase "*in God we trust, all others bring data.*". Such a catchphrase was particularly impactful at the time of its formulation, when few organizations had already implemented a sophisticated system of evidence-based metrics, but it is even more today, a historical era where data has visibly become the protagonist in most of the companies.

The purpose of this chapter is to supply a definition of big data and to provide its current organisational and management applications and possible joint directions with neuroscience. The present work, coming under the scope of an innovative process, with high levels of content originality, still appears to be unstructured and requires strengthened evidence both from academic and on-field sources. Additionally, limits and possible risks of this approach, including the privacy issue, will be addressed.

2. An overview on Big Data

Terminologically, the expression Big Data (BD) refers to a complex and voluminous quantity of information that is hard to process and analyse with traditional technology and therefore requires high-speed acquisition, archiving and analysis (Villars et al., 2011). For a more operational definition, the *5Vs model* is generally taken into account (eg,

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/

Demchenko et al., 2013; Hartmann et al., 2016), considering the following defining dimensions: volume, variety, velocity, veracity and value. Concerning the *volume*, a quantity of data is considerable "big" if greater than terabytes/petabytes (respectively 10^{12} and 10^{15} bytes, in the International System). The other parameters refer to the type of data (*variety*), generally not structured with a certain degree of multidimensionality, a high speed (*velocity*), intended as a measure of frequency of acquisition, storage and recovery. Finally, the data *veracity,* reflecting the level of reliability of the sources and the *value*, as the economic return due to the outcome, are also considered.

The transition from chaotic masses of data to high-value insights is achieved through differentiated processes, including data management, computer modelling, machine learning, inferential statistical (linear and non-linear) and information visualization techniques.

Through the BD approach, which relies on a very large amount of data, it is possible to identify patterns, correlations and latent structures that would not be observable with smaller samples, also allowing the studying of subpopulation or rare cases, with advantages in different fields: for example in epidemiology, where new insights with good statistical reliability are identified thanks to the clinical data derived from millions of hospital patients (Béjar-Prado et al., 2015).

The public sector is also increasingly making use of BD technology (Kim, Trimi & Chung, 2014). In fact, governments with the most advanced informatics infrastructures, have begun to implement it to increase efficiency, transparency and the security (both in terms of health and defence of citizens). One example among many is Israel, where the Ministry of Health has launched a project that uses BD and Artificial Intelligence (AI) for the localization of COVID-19 outbreaks through the administration of a questionnaire (lasting a minute on a daily basis) on citizens (Weizmann Institute, coronaisrael index, 2020).

Other applications are common in the clinical field, for example, delivering classification methods for electrocardiogram and automatic detection tools for arrhythmias (Kumar & Hannah Inbarani, 2015), which maintain the previous accuracy but can identify more categories.

The BD approach allows to unlock new perspectives and insights, revealing new opportunities. The reasons of its development can be ascribed to various factors. Firstly, the technological one, including: the reduction of architecture costs (hardware and software) and the data management processes (e.g., cloud computing), the distribution of sensors (e.g., IoT and BCI) and the increased computing power (also through the use of open source software such as Apache Hadoop and Python). A second factor, correlated to the first one, is the behavioural one, with the mass adoption of digital devices and the daily use of social media, allowing the collection of a never seen quantity of data, from heterogeneous sources, such as: IT (online transactions, browsing, social media, GPS tracking, sensors), financial, medical, biological. Finally, the third factor is the augmented complexity: the research, market and organizational phenomena are becoming

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/

increasingly complex and professionals need adequate techniques supporting their decisions and BD can, if used consciously, represent a useful tool.

3. Big Data in the organization

The interest in BD shown by corporates is justified by evidence indicating that organizations that use these technologies tend to perform better in terms of productivity and profit (Quaadgras et al., 2014).

One of the first steps for a company that wants to use BD is the mapping of data sources, namely those systems that produce information of any kind along the supply chain. Subsequently, the challenge, much more complex than the previous one, is to adopt (by purchasing or designing indoors, with ETL tools) an architecture (data management platform) and a data integration system for data of heterogeneous nature (e.g., tweets that are made up of a short, unstructured sequence of text or images and videos which are not optimized for semantic analysis), making sure that the process does not damage the quality of the information. Indeed, since quality (defined by the following dimensions: completeness, accuracy, type of format and value) (Setia et al., 2013) is specifically linked to the operational context, it is very important that during the integration operation it is preserved, as different sources can potentially compromise the outcome. Once integrated, data can be analysed (to name a few techniques: predictive analysis, data mining, language analysis) to obtain new insights and consequently make fact-based choices. Finally, the last step consists of the interpretation and communication of the results to a non-technical audience, in order to make the insights operational.

BD applications in the organizational field are various. Even though they are often ad-hoc solutions for specific company's problems, it is still possible to categorize them in three groups based on the aim: to strengthen internal processes (from marketing, to risk and human resources management), to enrichment of its products/services and customer experience and to monetize through the selling of data to external companies.

An example that represents a good case study in the marketing field is that of Uber (Cohen et al., 2016), a company that provides a private car transport service based on a mobile app. By using large masses of data, obtained through customers' smartphones, the company is able to implement a real-time pricing strategy, weighting supply and demand, market conditions and even road conditions (via API). The richness and density of the data collected (with detailed information for each session of the service, even when no transaction is made) lead to high economic returns.

Another example of BD application, although not exempt from criticisms (Angrave et al., 2016) and ethical issues (Illingworth, 2015), is in the personnel selection process (Hausknecht & Li, 2015; Landers & Schmidt, 2016), to the management of training and knowledge management, and to the reduction of staff turnover and to the evaluation by

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/

machine-learning of the psychological profile, of the performance and leadership potential, based on the information available from social network activities (Chamorro-Premuzic et al., 2017). An interesting case study is IBM one, where to reduce a high turnover and flight risk for strategic, an algorithm has been developed starting from the data available to the HR management (among others: performance, salary and promotions history, role, recruitment methods and also information about the sentiment) producing a decrease of 25% in the turnover in the following four years, a huge saving of resources and also an increased productivity (Corporate Research Forum, 2017).

4. Neuroscience and Big Data for the organization

As in the contexts previously described, even within neuroscience, the massive collection of data, when possible, could lead to great advantages (Kandel et al., 2013; Poldrack & Gorgolewski, 2014). However, it is crucial to understand the reasons.

Above all, it is known that congruent results obtained via different and complementary approaches [for example behavioral, electrophysiological (electroencephalography (EEG) or brain imaging, such as functional Magnetic Resonance Imaging (fMRI) or, functional Near-Infrared Spectroscopy (fNIRS)] tend to increase the degree of confidence and validity regarding a certain neurobiological phenomenon (Nichols et al., 2017). Secondly, research has shown that even in the neuroscientific field there is an increased statistical reliability and a greater predictive power (Costafreda, 2009). As with other sciences, neuroscience is also adapting to larger datasets (Poldrack &Gorgolewski, 2014), often obtained from the combined effort of institutes and research centers, that agglomerate data from hundreds and thousands of subjects. Despite this, the trend is not yet fully developed.

At the same time, there is an increasing need for advanced statistical techniques that go beyond the classical inferential ones that may prove unsuitable for complex multidimensional constructs. For example, deep learning techniques, which fall within the set of artificial intelligence (AI) tools, which use convolutional neural networks (CNN), have already been applied to EEG data, coming from many studies with different paradigms (Lawhern et al., 2018). Therefore, in the future, having a greater amount of data available, might lead to progress in the analysis of the EEG signal, possibly comparable to those occurring in the field of visual recognition and semantic analysis.

One of the fields that opens to wider potential successful scenarios and that might be enhanced by the intersection between BD and neuroscientific techniques is the organizational-managerial one. In fact, a deep understanding of cerebral-cognitive functioning on limited size samples, obtained by collecting evidence and matching it to the existing literature on neural models, can then be tested on large quantity of observational data, used an empirical test and model optimizer. This process can therefore

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/

lead to new developments, adding new tools and resources available to leaders assisting the management of HR functions such as training, rewards, talent supervision and promotions.

In this perspective, in an ongoing project of the International Research Center for Applied Cognitive Neuroscience (IrcCAN) at the Catholic University of the Sacred Heart of Milan, starting from the assumption that personality and psychological characteristics of an individual play an important role as predictors for the preference towards a visual stimulus, a research protocol has been developed in which, through the use of various perceptual factors, derived from the application of a machine learning script by Matz et al. (2019), we intend to develop a predictive algorithm that uses both neuroscientific metrics (obtained by electroencephalography, in particular events-related, autonomic and oculometric indices), and psychometric measures (Big Five Questionnaire), obtained during the exposure to visual stimuli during several trials.

In a second phase the algorithm will be tested on large numbers using social media data, to verify its predictive reliability. This outcome could have useful and important applications in the organizational field, for example, for the profiling of consumers with benefits for marketing and communication. In fact, the conjunction and diversification of the data source could add value and act as a solution, at least partial, to the limitations of BD techniques, a subject covered in the next section.

5. Caveats and limits

Since BD is a hot topic, it is good to scale back possible excessively high expectations and certain beliefs developed around the BD theme.

Firstly, it is important to remember that not all organizations need or can adopt this approach (due to entry barriers, skill gaps, different company objectives). In fact, as usual, the introduction of a tool must be considered and thought out also starting from the context of application and in case of implementation it is suggested to make it a gradual process.

Secondly, the power of these techniques does not substitute and cancel the need for human vision and intuition. Contrarily, there will be an increasing need for talent and skills: if the data becomes increasingly economic, the necessary skills to manage it are subject to the opposite trend, that is, they become increasingly rare and sought, also causing negative effects in countries that stayed behind (among other possible outcomes: unemployment, skill mismatch and increased inequality). Therefore, in a business context, it will be needed to attract and retain people with these skills by the talent management (perhaps using, as we have seen, also BD).

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/

Thirdly, it should be noted that even the power of BD is not absolute and free from critical and methodological issues. For example, one of them is the accuracy and quality of the data. Many of the studies that exploit BD are based on large datasets that were not created for such purposes (for example hospital datasets created for administrative aims): this certainly causes bias, the impossibility of carrying out longitudinal studies (because often the data is aggregated or processed with procedures that make it impossible to delete inaccurate or obsolete data) and critical issues in inferring cause-effect relations (mostly in the case of observational data).

Furthermore, another argument which reminds us the need of being cautious is the existence of situations presenting very worrying flaws, such as algorithms that discriminate certain social categories (Illingworth, 2015; Robinson et al., 2014). If it is true and worrying that the use of tools and techniques considered to be objective reassures and removes responsibility for the implementer, what damage can an instrument that contains discriminatory biases cause?

Finally, other potentially problematic issues are data ownership and the issue of privacy and data security as we will see in the next paragraph.

6. A MATTER OF PRIVACY

Lastly, it is appropriate to address the elephant in the room when it comes to BD, the issue of privacy and confidentiality, protected by laws, patents and contracts and strictly connected to data security. The subject becomes a thorny problem if the geographical factor is taken into account, namely the services of a certain company owning the data impact different legislations, and that regarding the type of data, under the assumption that greater security is needed for personal data (e.g., clinical) compared to browsing data.

However, data security and end-user protection are not always guaranteed. According to a survey on healthcare organizations (Ponemon Institute, 2014), 94% of respondents had a data breach in the previous two years, and on average, only 6% of the total annual IT budget was spent on cyber-security (HIMSS Media, 2016). The situation is aggravated knowing that BD predictive analysis are shown to have an important impact on people's well-being and careers (Newell & Marabelli, 2015).

To formulate a linear and effortless solution is clearly impossible. A first attention, perhaps obvious, is that of individual awareness. As an end-user, it is important to comprehend the basic mechanism of digital security, in order to avoid naïve attitudes, such as the exchange of sensitive information on unsecure platforms. Instead, at a more inter-individual level, it will be mandatory to make organizations invest a consistent part of the budget on data security and adopt standard processes that always enhance more privacy protection.

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/

In this direction, the General Data Protection Regulation (GDPR), effective as of May 2018, was designed. Among many updates, pragmatical indication of behaviour were given, such as: how and when report a data breach, pseudonymisation and encryption of data, the need to restore access to personal data in the event of system error, and the introduction of testing procedures to ensure security.

Far more ambitious, maybe utopian, is another concept introduced by the GDPR: *the right to be forgotten*. In a now historic ruling (C-131/12, 2014), the European Court of Justice has imposed the removal of any content that may be "incorrect or potentially damaging to the private sphere and image" on Google Inc, in the face of the request of the interested party, in the name of mentioned right. Google seems to have assimilated this evolution, subsequently inserting an online form that allows the request for content removal. However, there is an obvious need to develop techniques that automate this process and are scalable to the size of the web. The development of such a function represents a possibility, even if it presents problems such as the difficulty in establishing whether a certain content is of public interest or not and what are the elements that justify the persistence of data in the archives, even after some time high.

7. Conclusions and future challenges

BD technology is a complex phenomenon, with great potential in many sectors, still to be fully exploited in some cases. In the next future for organizations, an even more intense and interdisciplinary application of it is expected, with tailor-made tools that will provide value due to the development of new technologies (e.g., AI, 5G)

At this moment we are therefore called to question ourselves and the impact of these new technologies, guaranteeing the purse of economic objectives while considering, step by step, the well-being of people and the ethical dimension. In the meantime, institutional bodies play a key role, being responsible for protecting end-users, without damaging the economic interests of companies. It is necessary to strengthen communication between policymakers, professionals and citizens, fully embracing the idea of positive technology. As perfectly formulated by writer Christian Lange: "*technology is a useful servant but a dangerous master*".

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/

References

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal, 26*(1), 1-11. doi: 10.1111/1748-8583.12090

Béjar-Prado, L., Gili-Ortiz, E., & López-Méndez, J. (2015). Exploitation of Healthcare Databases in Anesthesiology and Surgical Care for Comparing Comorbidity Indexes in Cholecystectomized Patients. In J. Kacprzyk, J. Abawajy, A. Taher Azar, V. Snasel, A. E. Hassanien (Eds.) *Big Data in Complex Systems*, (263-292). Cham, Springer. doi: 10.1007/978-3-319-11056-1_9

Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., & Sherman, R. A. (2017). The datafication of talent: How technology is advancing the science of human potential at work. *Current Opinion in Behavioral Sciences, 18*(1), 13-16. doi: 10.1016/j.cobeha.2017.04.007

Cohen, P., Hahn, R., Hall, J., Levitt, S., & Metcalfe, R. (2016). Using big data to estimate consumer surplus: The case of uber (No. w22627). *National Bureau of Economic Research*. https://www.nber.org/papers/w22627.pdf. doi: 10.3386/w22627

Costafreda, S. G. (2009). Pooling FMRI data: meta-analysis, mega-analysis and multi-center studies. *Frontiers in Neuroinformatics, 3* (33), 1-8. doi: 10.3389/neuro.11.033.2009

Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. 2013 IEEE International Conference on Collaboration Technologies and Systems, 48-55. doi: 10.1109/cts.2013.6567203

European Parliament and Council of European Union (2016) Regulation (EU) 2016/679. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data–a taxonomy of data-driven business models used by start-up firms. *International Journal of Operations & Production Management. 36*(10) 1382-1406. doi:10.1108/ijopm-02-2014-0098

Hausknecht, J. P., & Li, H. (2015). Big data in turnover and retention. In S. Tonidandel, E. King, & J. Cortina (Eds.), *Big data at work: The data science revolution and organizational psychology* (pp. 250 –271). NewYork, NY: Routledge

Illingworth, A. J. (2015). Big data in IO psychology: Privacy considerations and discriminatory algorithms. *Industrial and Organizational Psychology*, *8*(4), 567-575. doi: 10.1017/iop.2015.85

Kandel, E. R., Markram, H., Matthews, P. M., Yuste, R., & Koch, C. (2013). Neuroscience thinks big (and collaboratively). *Nature Reviews Neuroscience*, *14*(9), 659-664. doi: 10.1038/nrn3578

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78-85. doi: 10.1145/2500873

Kumar, S. S., & Inbarani, H. H. (2015). Modified soft rough set based ECG signal classification for cardiac arrhythmias. In J. Kacprzyk, J. Abawajy, A. Taher Azar, V. Snasel, A. E. Hassanien (Eds.) *Big Data in Complex Systems*, Cham: Springer, pp. 445-470. doi:10.1007/978-3-319-11056-1_16

Landers, R. N., & Schmidt, G. B. (2016). Social media in employee selection and recruitment: An overview. In *Social Media in Employee Selection and Recruitment*, Cham: Springer, pp. 3-11. Springer, Cham. doi: 10.1007/978-3-319-29989-1_1

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *15*(5), 140-151. doi:10.1088/1741-2552/aace8c

Matz, S. C., Segalin, C., Stillwell, D., Müller, S. R., & Bos, M. W. (2019). Predicting the personal appeal of marketing images using computational methods. *Journal of Consumer Psychology, 29*(3), 370-390. doi: 10.1002/jcpy.1092

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. *The Journal of Strategic Information Systems*, 24(1), 3-14. doi: 10.1016/j.jsis.2015.02.001

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., ... & Proal, E. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature neuroscience*, *20*(3), 299-303. doi: 10.1038/nn.4500

Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: data sharing in neuroimaging. *Nature neuroscience*, *17*(11), 1510-1517. doi: 10.1038/nn.3818

Ponemon Institute (2015) Fifth Annual Benchmark Study on Privacy and Security of Healthcare Data. http://www.ponemon.org/blog/criminal-attacks-the-new-leading-cause-of-data-breach-in-healthcare

Quaadgras, A., Weill, P., & Ross, J. W. (2014). Management commitments that maximize business impact from IT. *Journal of Information Technology*, *29*(2), 114-127. doi:10.1057/jit.2014.7

Report Cybersecurity, eHealth trendbarometer. HIMSS Media (2016). https://www.himssmedia.com/

Robinson, D., Yu, H., & Rieke, A. (2014, September). Civil rights, big data, and our algorithmic future. In Leadership Conference on Civil and Human Rights. https://bigdata. fairness. io/wp-content/uploads/2014/11/Civil_Rights_big_Data_and_Our_Algorithmic-Future_v1 (Vol. 1).

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/

Setia, P., Venkatesh, V., & Joglekar, S. (2013). Leveraging digital technologies: How information quality leads to localized capabilities and customer service performance. *Mis Quarterly*, *37*(2), 565-590. doi: 10.25300/misq/2013/37.2.11

Strategic Workforce Analytics. Corporate Research forum: IBM (2017). https://www.orgvue.com/resources/research-report/strategic-workforce-analytics-research-report/

Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big data: What it is and why you should care. White paper, IDC, 14, 1-14. http://www.tracemyflows.com/uploads/big_data/idc_amd_big_data_whitepaper.pdf

Weizmann Institute, coronaisrael index (2020). https://coronaisrael.org

M. Balconi (ed.), *Neuromanagement: People and Organizations*, Milano, LED, 2020
ISBN 978-88-7916-952-3 - https://www.ledonline.it/IrcCAN-Cognitive-applied-neuroscience/